

# Application of a Mixed Effects Model for Biosurveilliance of Regional Rail Systems – Illustration with the PATCO Regional Rail Line

Robert J. Gallop  
West Chester University  
7 February 2007

Quantitative Methods in Defense and National Security

# CiMerC

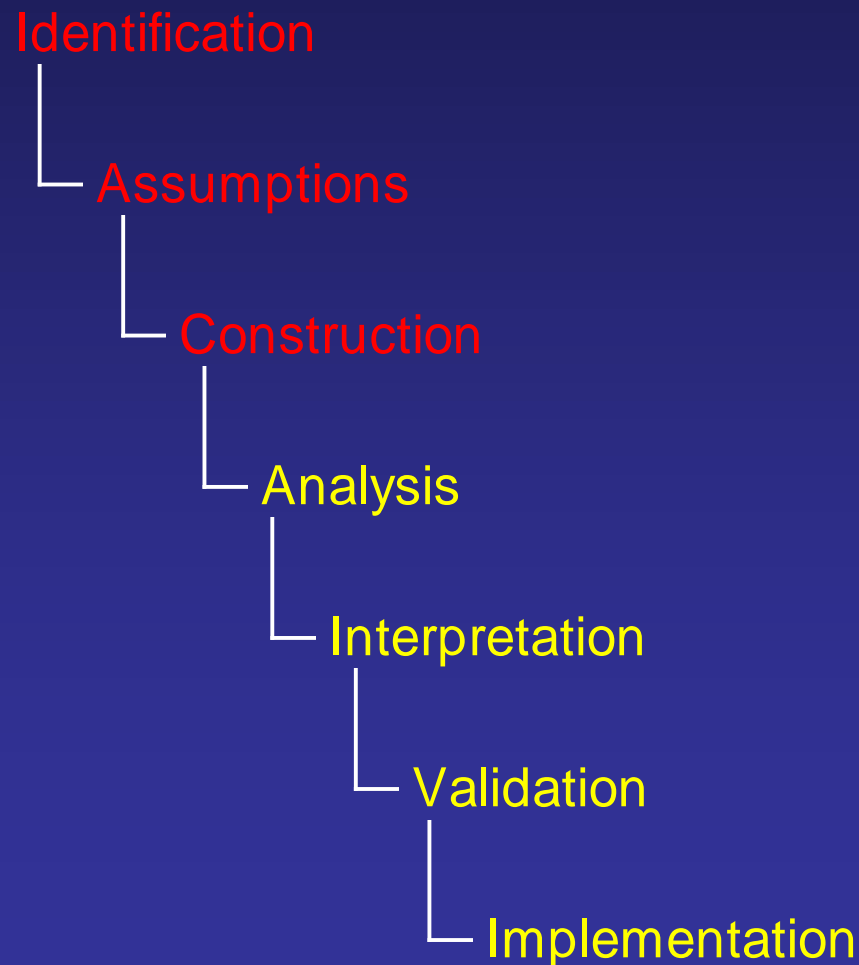
- National Bioterrorism Civilian Medical Response Center
- <http://www.cimerc.org>

# Aims and Objectives

1. Briefly Describe the Biosurveillance settings
2. Describe the statistical process and available data
3. Illustrate the Application of the Mixed Effects Model
4. Discuss limitations and extensions

Evans, S., Kleinman, K., and Pagano, M. (2002). Bioterrorism and Biosurveillance, *AMSTAT News*, 301, 41-43.

# The process of Modeling can be broken up into seven stages



# Need for Biosurveillance









**Focus on Surveillance in a subtle attack**



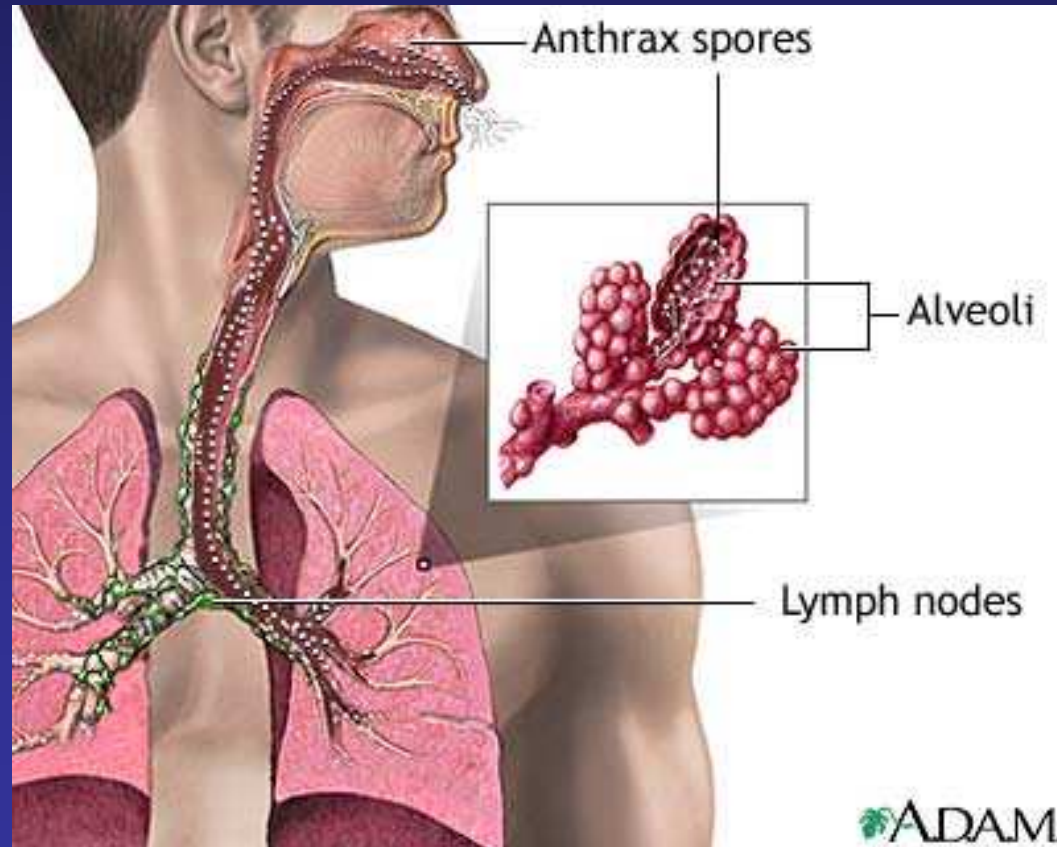
**The release of a Chemical/Biological agent**

# History

- 1754 – Sir Jeffrey Amherst gives American Indians blankets and handkerchiefs containing smallpox.
- 1917 – Germany attempts to infect livestock destined for shipments to the U.S. and Russia with anthrax and glanders.
- 1979 – Outbreak of inhalational anthrax in Russia, 79 cases, 68 deaths.
- 2001 – 10 confirmed cases of inhalational anthrax in D.C., Florida, and New York due to mailed letters/packages containing *B. anthracis* spores.

# Subtle attack: Inhalational Anthrax

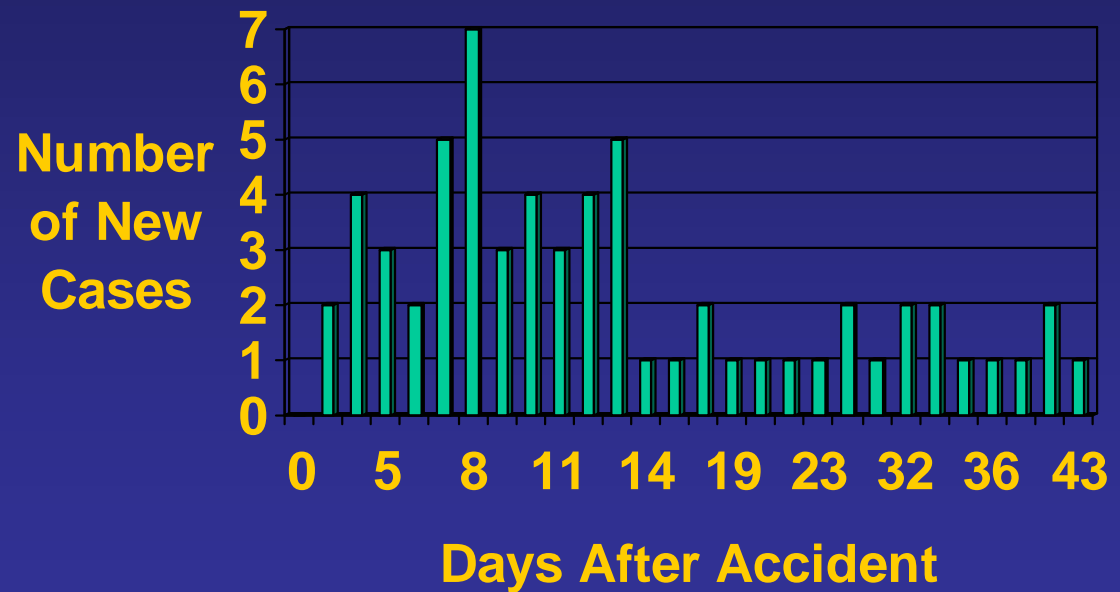
- Cause: Breathing in "aerosolized" anthrax spores



# Symptoms

- Not Immediate – Incubation Period

## Russian Inhalational Anthrax incident



# Symptoms continued

- Early Phase: Fever, Cough, Chest Pains, Nausea, Vomiting, Abdominal Pains
- Could experience a brief period of apparent recovery
- Late Phase: Abrupt rise in fever, dyspnea and shock, Hemorrhagic mediastinitis, CXR with widened mediastinum, Meningitis in 50%, 90% mortality (40% in recent attacks)

# Identification:

## Scenario for an Attack involving Regional Rail System

- **Key location:** Transportation hub at Philadelphia's 8th Street and Market Street station.
- **Method:** Through ventilation systems of a targeted hub / transportation mode
- **Result 1:** Development of Inhalational Anthrax after incubation period.
- **Result 2:** Upon presentation of the infection as inhalational anthrax, many commuters may be too sick to go to work.
- **Result 3:** Many infected individuals will not attribute the infection to bioterrorism

# Biosurveillance – Monitoring Number of commuters

- Monitoring the passenger loads for regional rail systems through a mixed effects model approach.
- This mixed effects model approach will allow us to forecast the expected number of commuters on a regional rail system.
- If the observed number is well below the expected number this raises possible suspicions.

# Available Data – Patco Data

- Data from the Port Authority Transportation Company (PATCO) rail system, which serves New Jersey residents who commute to Philadelphia and Philadelphians who commute to New Jersey

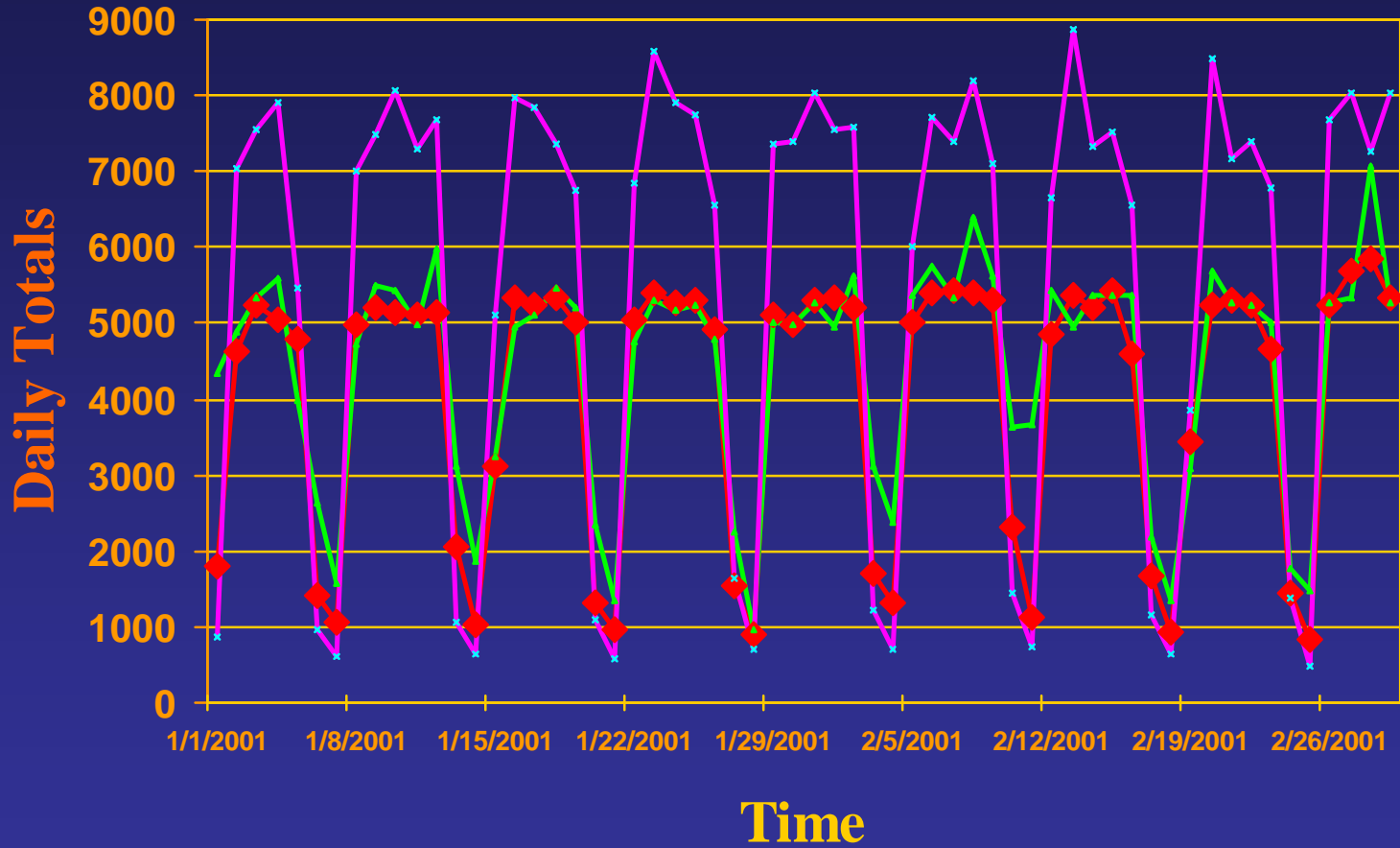




# Data Description

- We have data for each station in New Jersey and Philadelphia for 2001.
- The data consists of daily total counts of the number of passengers entering each station
- The PATCO data consists of 13 census tracts
- Consisting of 365 data points for each tract
- Corresponding to the daily total per station for the 2001 calendar year

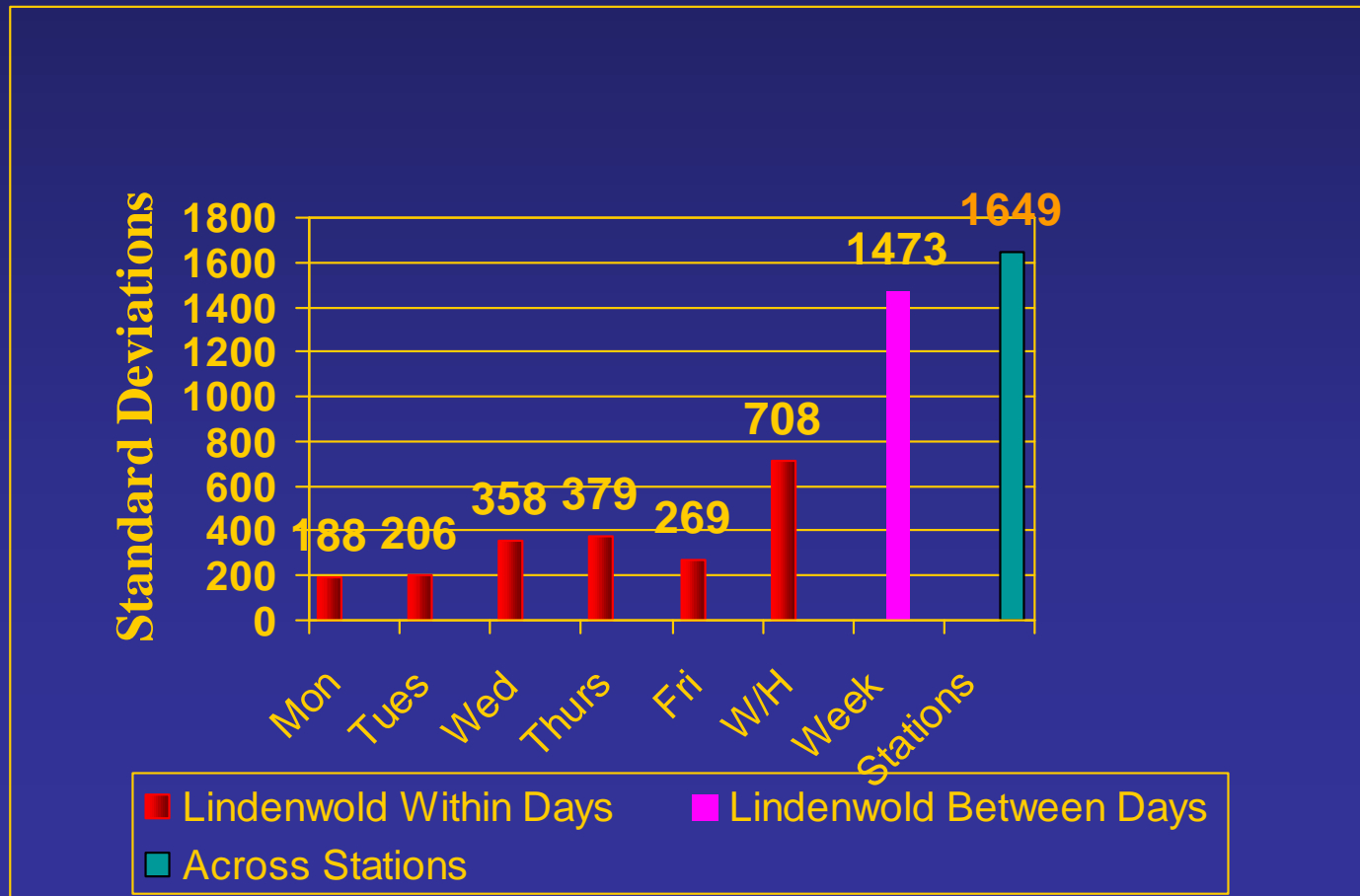
# Data Illustration



◆ Lindenwold    — 8th Street    —x— 16th Street

# Sources of Variability

- Variability within the various Monday daily totals as well as for the other days (within day classification per station source of variability )
- Within a station, there is variability across the days of the week
- Variability across stations



# Assumptions and Construction: Mixed Effects Model

- For the PATCO data, the hierarchy of the data consists of three levels:
- Reported observation for each day of the week and station combination
- Day of the week classification within station
- Station.

# Mixed Effects Model (continued)

$$Y_{ijk} = \beta \text{ (grand average)} + \varepsilon_k + \varepsilon_{j(k)} + \varepsilon_{i(j(k))}$$

$Y_{ijk}$  is the  $i^{\text{th}}$  observation for the  $j^{\text{th}}$  day classification of the  $k^{\text{th}}$  station,

$\varepsilon_{i(j(k))}$  accounts for the random variance within a day classification,

$\varepsilon_{j(k)}$  accounts for the random day classification variance nested within station, and

$\varepsilon_k$  accounts for the random station variance.

Note for  $Y_{ijk}$ ,  $i = 1, 2, \dots, 52$  corresponding to the 52 weeks of the year,  $j = 1, 2, \dots, 6$

corresponding to the day of the week classifications, and  $k = 1, 2, \dots, 13$  corresponding to the 13 rail stations on the PATCO line.

The assumptions for the random elements are:

$$\varepsilon_{i(j(k))} \sim \text{NIID}(0, \sigma_e^2)$$

$$\varepsilon_{j(k)} \sim \text{NIID}(0, \sigma_{DS}^2)$$

$$\varepsilon_k \sim \text{NIID}(0, \sigma_s^2).$$

The variance of  $Y_{ijk}$  is:

$$\text{Var}(Y_{ijk}) = \sigma_s^2 + \sigma_{DS}^2 + \sigma_e^2.$$

# Expected Passenger Loads

- Each station's passenger load is characterized by a set of station and day classification random effects.
- Estimates are termed Best Linear Unbiased Prediction (BLUP) estimates.
- BLUP estimates are linear in the sense that they are linear functions of the data; unbiased in the sense that the average value of the estimates is equal to the average value of the quantity being estimated; best in the sense that they have the minimum mean squared error within the class of linear unbiased estimators; and prediction estimates to distinguish them from estimation of the random effects

# Threshold of Concern

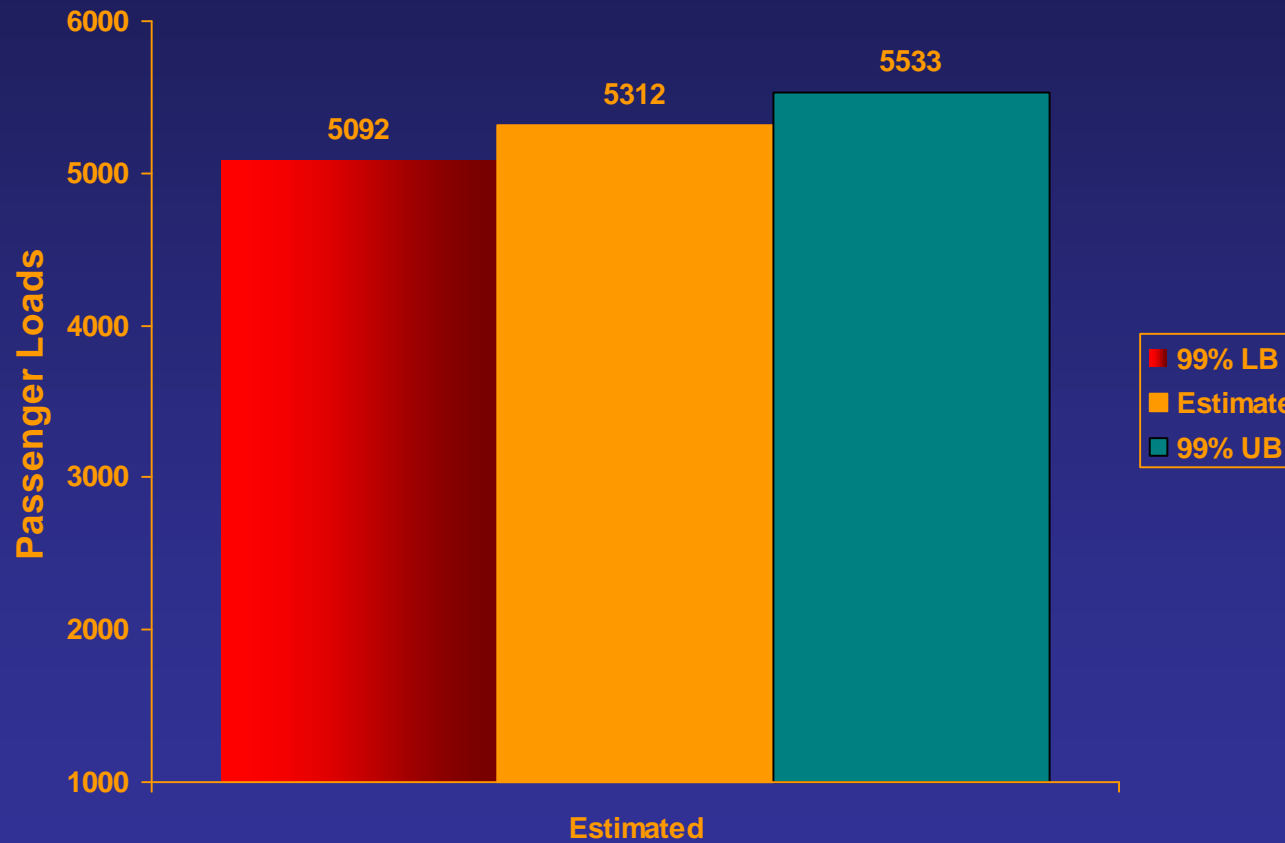
- We establish an acceptable reduction of less than 35% reduction in average estimated number of passengers for a station on a given day.



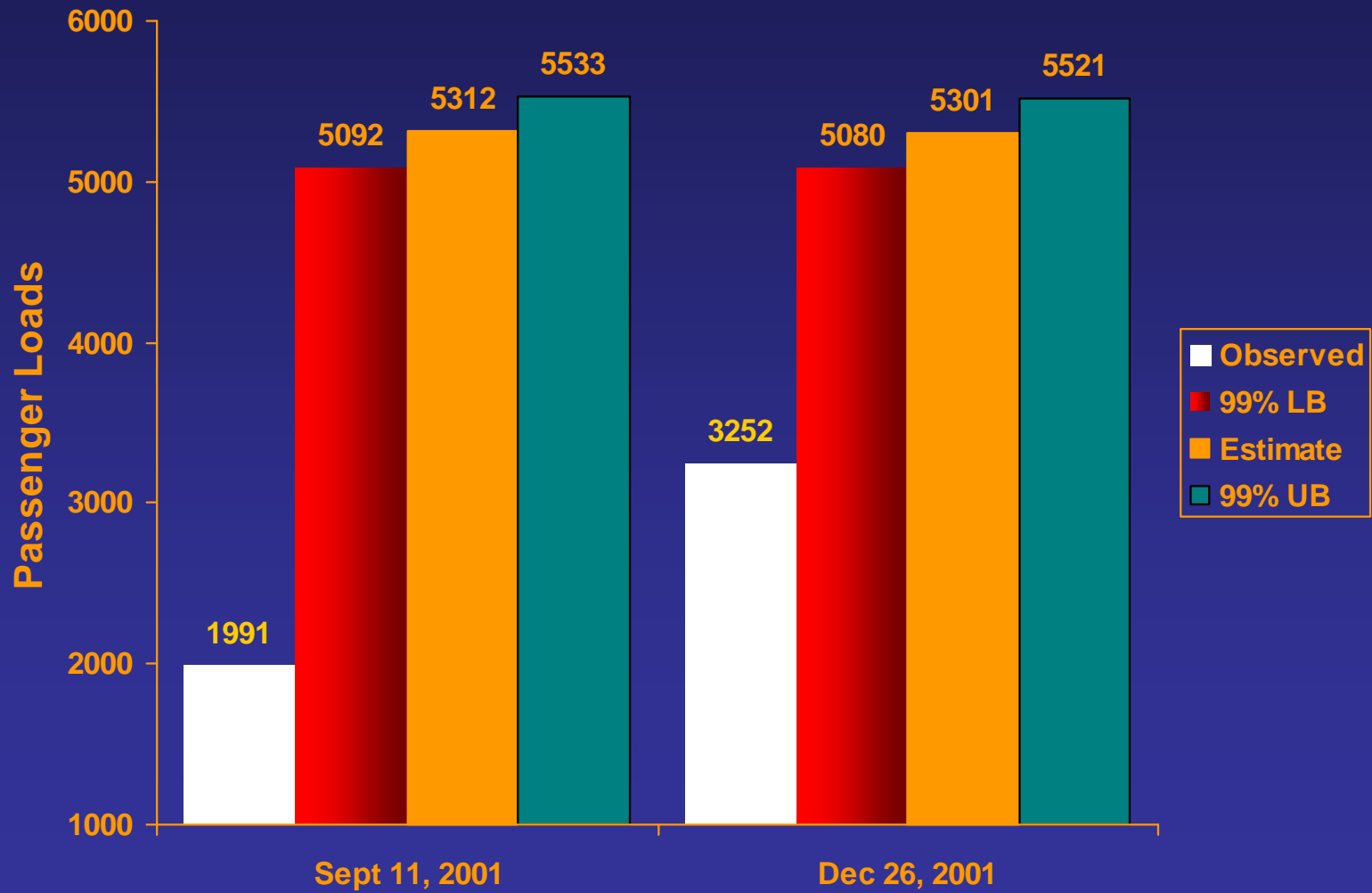
# Analysis and Interpretation: Biosurveillance Application

- We split the 2001 PATCO data in two halves.
- We use the first six months to estimate expected daily totals for each station on the PATCO line
- We will compare these estimates to the observed daily totals in the second six months.

# Estimated Passenger Load – Phila 8<sup>th</sup> and Market for Tuesday



# Two Occurrences



# Future Implementation: Real-time setting - Data

- PATCO regional rail line discussed in this example has integrated magnetic / smart card fare system
- Daily Customers are no longer generic market segments. PATCO has the ability to know exactly who they are, when they typically use the PATCO facilities, and what is happening at the current moment.

# Real time setting - Model

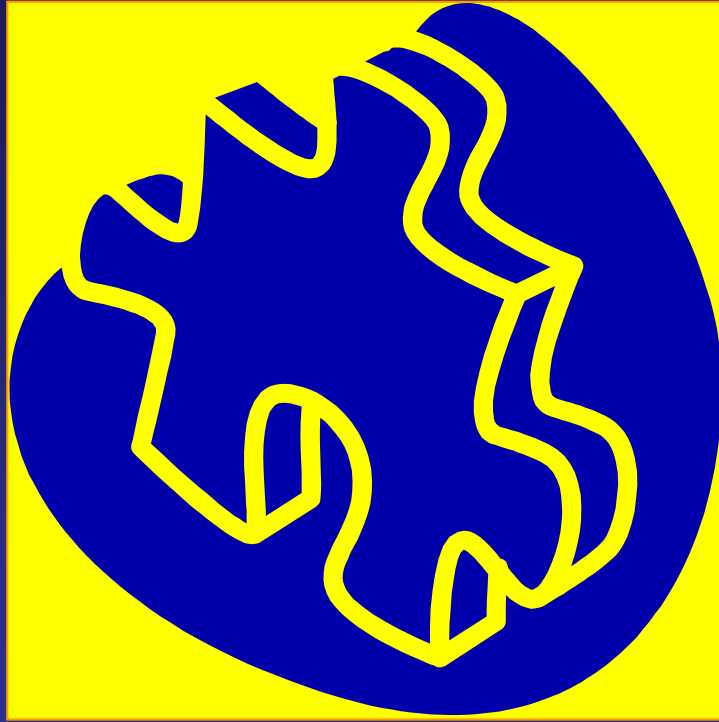
- SAS/IntrNet Application Dispatcher is a compute service that allows users to pass parameter selections from a Web page to the appropriated SAS program that executes on a server, sending the results back to the user's Web browser.
- Because the program is executed on the server, end users only need a Web browser. There is no need for SAS on their workstations.
- The Web browser can be automated to refresh itself periodically.
- Combining the current database with SAS/IntrNet Application Dispatcher, reports will be generated upon a specific scheduled train conclusion of its inbound or outbound commute

# Conclusions

- We have presented an application of mixed effects model for estimation of passenger loads on the PATCO regional rail line for any day or station
- This mixed effects model has three benefits. First, it accommodates the hierarchical structure of the data, which has repeated observations per day classification nested in station. Second, it provides best estimates of the expected number of passengers per station for any day through the derivation of BLUP estimates. Third, there are a vast number of statistical packages to implement this model

# Limitations

- Distribution theory associated with BLUP estimates is not nearly as well-understood as it is with conventional estimable functions
- The variance of the BLUP estimates may experience shrinkage, since the observed data are shrunk towards the overall average since the prior means of the random effects is zero
- Acceptability of the derived threshold rule









# References

- Gallop, R.J., Mode, C.J., Blank, K., Jurgens, S.M., and Schaben, C. (2005). Application of a Mixed Effects Model for Biosurveillance of Regional Rail Systems, *Journal of Data Science*, 3, 353-370.
- Banks, D. (2002). Statistical Issues in Counter-Terrorism, Retrieved 10 October 2002, <http://www.statdns.org>
- Brannon, W.P. (2003). Regional Fare Collection and Customer Data Management, Retrieved 14 January 2004, [http://spider.apta.com/gwf/transitech/2003/brannon\\_w\\_price.pdf](http://spider.apta.com/gwf/transitech/2003/brannon_w_price.pdf)
- Brannon, P. (2003). Biosurveillance, Regional Meeting on NBC procedures, Philadelphia, PA.
- Evans, S., Kleinman, K., and Pagano, M. (2002). Bioterrorism and Biosurveillance, *AMSTAT News*, 301, 41-43.
- Robinson, G.K. (1991). That BLUP is a good thing, *Statistical Science*, 6, 15-51.
- Timbers, V. (2003). Connecting the SAS System to the Web: An Introduction to SAS/IntrNET Application Dispatcher, *Proceedings from the 2003 North Eastern SAS User's group Conference*, Cary, NC: SAS Institute Inc.